International Scholars Journals

African Journal of Agronomy ISSN: 2375-1185 Vol. 10 (9), pp. 001-006, September, 2024. Available online at www.internationalscholarsjournals.org © International Scholars Journals

Full Length Research Paper

Non-Destructive Detection of Moldy Chestnuts via Near Infrared Spectroscopy

J. Liu[1], X. Y. Li[1]*, P. W. Li[2], W. Wang [1], J. Zhang[1,3], R. Zhang[1] and P. Liu[1]

[1]College of Engineering, Huazhong Agricultural University, Wuhan, 430070, China. [2]Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan, 430062, China. [3]Key Laboratory for Highway Construction Technology and Equipment of Ministry of Education, Chang'an University, Xi'an, 710064, China.

Accepted 16 August, 2024

One of the key agricultural crops, particularly in Asia, is chestnut. However, following harvest, mildew causes a significant loss of chestnuts. Finding and removing mouldy chestnuts is a crucial step in preventing and minimising widespread mildew. Additionally, this method is crucial in the manufacturing of food products made from chestnuts. In this study, we used near infrared spectroscopy to pinpoint mouldy chestnuts. A discriminating model was built using the near infrared spectra of 109 samples of chestnuts, including 40 samples without mildew, 40 samples with severe mildew, and 29 samples with mild mildew, spanning the wavelength range of 833 to 2500 nm. For the model's validation, a second set of samples with three mixed groups—chestnuts without mildew (n = 20), chestnuts with severe mildew (n = 20), and chestnuts with light mildew (n = 8)—was employed. The findings demonstrate that by utilising the first derivative and vector normalisation for spectra preprocessing and the Ward's algorithm as a distance algorithm approach, the best classification model based on the spectral band of 1818 - 2085 nm was attained. Sound chestnuts, somewhat mouldy chestnuts, and severely mouldy chestnuts were correctly classified at rates of 100, 92.8, and 100%, respectively. These findings showed that the near infrared spectral analysis-based discriminating model can successfully identify rotten chestnuts.

Key words: NIR, supervised pattern recognition, nondestructive detection, chestnuts, mildew.

INTRODUCTION

As an important agricultural product, the chestnut is popular for its delicious flavor and abundant nutrition in Asia and Europe. More than 11,000,000 ton chestnuts were consumed annually all over the world (http://faostat.fao.org/site/336/DesktopDefault.aspx?Page ID = 336, 2007). However, chestnuts tend to develop mildew after harvest for its abundance of alimentation and moisture which leads to storage lost and the degradation of final products. The most efficacious and safe method for reducing and preventing large-scale chestnuts mildew is to identify and remove the moldy chestnuts. In addition, it is also important to eliminate moldy chestnuts to ensure the quality of processed chestnut products such as powder, can, confection, juice and other category (Freinkel, 2009).

However, the presence of the peel makes it difficult to detect or assess the condition of the nut. In current production, the dominating methods of eliminating the degenerative chestnuts, are removing the floating mold nuts from salt solution or manually sorting out the appearance-changed ones, according to the visual observation of the appearance. Both of them are time-consuming, labor-intensive and incapable of classifying the slight molds nut without obvious surface difference from the sound ones. If the peel were removed, the utilization would be constrained because more than half of the chestnuts are marketed as whole chestnut products with the peel being intact in Asia. Furthermore, it also adds considerable cost by storing peeled nut kernel (Feng, 2007). Thus, it is necessary to develop a rapid nondestructive technique for detection of moldy chestnuts with peel,

*Corresponding author. E-mail: lixiaoyu@mail.hzau.edu.cn. Tel: +86 27 87282120. Fax: +86 27 87282121.

Table 1. Chestnut sample groupings for near-infrared classification.

Chestnut sample	Sound chestnut (G)	Slightly moldy (P)	Severely moldy (S)	Total
Calibration set	40	29	40	109
Validation set	20	8	20	48
Total	60	37	60	157

which would greatly benefit growers, processors and customers.

Near infrared (NIR) spectroscopy is a nondestructive measuring method with the advantages of minimized preparation of samples, fast and easy to operate and environment-friendly (Lu et al., 2007). It has been conducted to do both quantity and quality analysis of agricultural products, (Yan et al., 2007, Nicolai, et al., 2007, Zhu, et al., 2009). The application of NIR in the classification analysis has multiple targets including, quality level of wheat flour (Cocchi et al., 2005), grade of Chinese traditional herb (Wang et al., 2007), geographic origin of green tea (Chen et al., 2009), brands of instant noodles (Liu et al., 2008) etc. It was also used to estimate the mold in food (Di Egidio et al., 2009; Tripathi et al., 2009), inspect the infestation of fig (Burks et al., 2000), and detect the insect in grain (Neethirajan et al., 2007). For the targets with peel, NIR still is an effective means for measuring interior properties since the penetration of NIR is more than 9 mm.

It has been successfully implemented in quantitative analysis of targets with peel, such as oranges (Tewari et al., 2008) avocados (Blakey et al., 2009), peanuts (Sundaram et al., 2009) and intact asparaguses (Flores-Rojas et al., 2009). Hence, NIR has the potential for measuring the properties or condition of the chestnut kernel inside of the peel with reduced time and labor compared to conventional methods, which would benefit the growers and consumers.

In this research, we explored the application of NIR spectroscopy in measuring the target with peel, evaluated the feasibility of NIR spectroscopy in measuring the kernel properties of chestnuts, compared the effect of different bands, distance algorithms and spectra preprocessing methods in the model calculation respectively, and established a relationship between the spectra and the mold extent of chestnut.

MATERIALS AND METHODS

Sample preparations

Chestnuts for this experiment were purchased from a local market. Half of them were stored under standard refrigeration conditions (0 - 4°C and 80 - 90% relative humidity) while the other half of them were kept in natural conditions (17 - 22°C and 40 - 60% relative humidity). These conditions are commonly applied to store chestnuts during which mildew occurs. All the samples were mechanically opened after spectra acquirement, in order to observe the interior conditions. Three groups of samples were collected: sound chestnut samples, labeled as group G; slight moldy chestnuts samples, labeled as group P; and severe moldy chestnut samples, labeled as group S.

Chestnuts without moldy kernels were classified as the sound chestnut samples while those with more than 3/4 moldy kernels, were considered as the severe moldy chestnut samples. The chestnuts with less than 3/4 moldy kernels were classified to be the slightly moldy chestnut samples. The reason for distinguishing the moldy chestnuts into two groups is that, the presence of mildew in chestnut is a developing process. Mildew normally starts from a small region kernel and then proliferates to all the regions of the kernel. No noticeable difference was observed from the appearances of the three sample groups. Table 1 summarizes the sample information for each group for calibration and validation.

NIR spectral measurement

NIR spectra were acquired using a VECTOR33 NIR spectrometer (VECTOR33, Brucker Optic GMBH,, Germany) configured by fiber optic accessories. The aperture setting of the instrument is 1.4 mm and the beam splitter is quartz. Each spectrum has 2075 data points, as the result of 64 scans among the wavelength range from 833 to 2500 nm. The spectrum of each sample was acquired using the diffuse reflectance mode and absorbance was calculated for each spectrum. The environment temperature was kept around 26°C while the humidity was kept at an ambient level in the laboratory.

The samples were exposed to temperature 26°C for 2 h before experimental tests to allow for temperature equilibrium. The spectrometer was warmed up for 1 h prior to the spectra collection. The acquired spectra were stored on a personal computer connected via an AQP card to the spectrometer, using the software OPUS (Brucker Optic GMBH, Germany, 2004). The intact chestnut samples were scanned with their flat side facing the probe. After spectra acquirement, each chestnut sample was peeled off and cut open to identify the presence of mold in the chestnut and the degree of the mold.

Preprocessing methods

Preprocessing of raw spectra is necessary in NIR spectroscopy modeling, in order to eliminate the interference caused by random noise, baseline drift, temperature and other factors. Different preprocessing methods may have different effects on the model development and it is thus, important to select an appropriate preprocessing method in developing a NIR calibration model. Vector normalization (VN), first derivative (FD), first derivative and vector normalization (FD and VN), second derivative (SD) and second derivative and vector normalization (SD and VN) are commonly used preprocessing methods.

These methods plus no preprocessing (NP) were compared, to find the most accurate preprocessing methods.

Figure 1. The photographs of three group samples: a) sound chestnut sample, b) slightly moldy chestnut sample and c) severely moldy chestnut sample.

Calibration and validation of the models

Classification models were established based on the data derived from different sample groups, using a supervised pattern recognition method. Supervised pattern recognition has been applied in biology, chemistry and pharmaceutical and has been explored for food identification combined with NIR spectroscopy (Roggo, 2003). With this method, prior knowledge of the samples belonging to the different categories, is used in the classification procedure. Since the performance of classification algorithm depends on the data structure and the results from different distance algorithm methods are different, even based on the same sample data, the main work in application of supervised pattern recognition is to choose the most appropriate one from numerous methods. The classification effects of seven distance algorithm methods were compared, to ascertain the most accurate method for classification model. The seven distance algorithm methods were single linkage (SL), complete linkage (CL), average linkage (AL), weighted average Linkage (WAL), median algorithm (MA), centroid algorithm (CA) and Ward's algorithm (WA).

The classification models were established using the data of different groups of samples and their predictions were accomplished by comparing the information of target samples with the prior knowledge of the possible categories which the target sample would belong to (Chen, 2009). The effectiveness of the models were evaluated using an independent validation set that has not been used in the model development.

A set of spectra for the three chestnut groups (that is, sound chestnuts without mildew, slight mildew and severe mildew) were selected as the training set and the distances between the spectra were calculated to function, as a standard for screening moldy chestnuts without knowing the mildew conditions. When a spectrum needs to be identified, the distance between this spectrum and the training set was computed and the judgment would be given according to this value. The models based on different preprocessing methods, distance algorithm methods and spectrum bands were compared, to find conditions for establishing the optimized model.

The classification model was validated by analyzing and comparing the predicted mildew conditions of the chestnuts with the actual mildew conditions of the same samples by visual inspection. The effectiveness of the model, was evaluated according to the correct classification rate of the independent validation set. All analyses were carried out by using the software OPUS (Brucker Optic GMBH, 2004).

RESULTS AND DISCUSSION

Three sample groups and original spectra

Figure 1 shows the representative sound, slightly moldy and severe moldy chestnut samples. There was no noticeable difference between the outside appearances of the three sample groups. However, after opening, it could be found that the mildew occurred more than 3/4 of the fruit, on the severely moldy chestnut sample (c). The slightly moldy sample (b) only had a small area of mildew in the lower-right part while the other part of the fruit had the same appearance with the sound sample (a).

Figure 2 shows the representative spectra of the three chestnut sample groups. For the entire spectral region, the sound chestnuts had relatively lower absorbance while severely moldy chestnuts showed the highest absorbance compared with the other two sample groups. This resulted from the decomposition of the fruit which led to the change of the reflection properties. The spectra of slightly moldy chestnuts generally fell into a range between the group G and group S, in the spectral region of approximately 1,400 - 2,400 nm. The difference of the absorbance was consistent with the progress of the mildew: the components in a slightly moldy sample varied more than a sound one and less than a severely moldy one. They also presented similar configurations with group G in the region of 1,200 - 1,900 nm. The three sample groups exhibited more pronounced differences for the spectral region of 1,900 to 2,400 nm. However, it was still difficult to distinguish the three groups of samples because of the overlapping and similarity of the spectra in this region.

Comparison of the preprocessing methods and distance arithmetic

Table 2 shows results from models established by different preprocessing and distance algorithm methods, based on the calibration set of G and S sample spectra in the entire spectra region (833 - 2500 nm). It was found that the average correct classification rates of the models, based on the seven distance algorithms were 47, 62.3, 76.8, 67.5, 59.7, 71.2 and 85.8%, respectively. Therefore, the Ward's algorithm was the most accurate and effective method for classification model. The definition of this algorithm which uses the increase in variance for the cluster being merged, as the criterion would contribute to its power of distinguishing the spectra of chestnut samples. Furthermore, several preprocessing methods, including second derivative and vector normalization, second derivative, first derivative and vector normalization and no preprocessing methods, could be successfully used with the Ward's algorithm for accurate classification

of moldy chestnuts, since the average correct classification rate of these models are 96, 95, 92

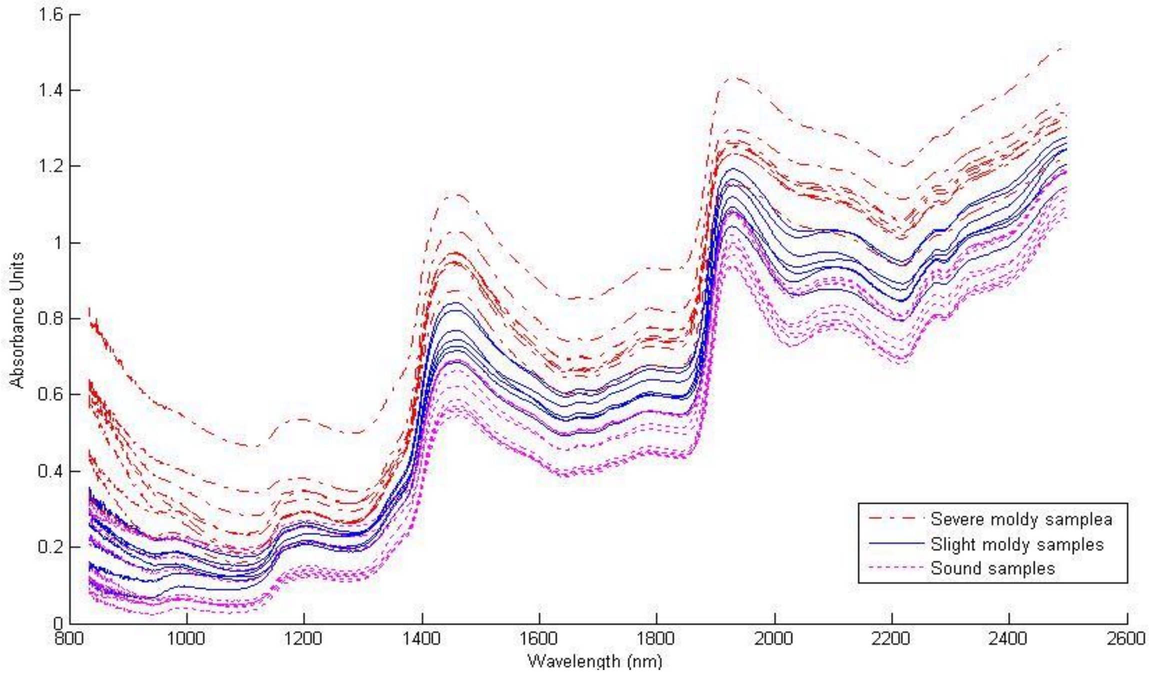

Figure 2. The typical NIR spectral of three kinds of samples.

Table 2. Calibration results based on different preprocessing and distance arithmetic methods.

	Correct classification rate (%)											
	NP		VN		FD		FD+VN		SD		SD+VN	
	G	S	G	S	G	S	G	S	G	S	G	S
SL	98	30	100	12	100	2	100	6	100	10	90	6
CL	88	76	100	22	98	28	90	6	72	24	96	48
A L	88	80	100	22	98	26	88	68	98	80	94	80
WAL	88	90	100	22	98	26	94	74	96	16	100	6
M A	88	96	100	8	98	2	100	10	100	10	100	4
CA	90	98	98	24	90	86	90	64	100	10	100	4
WA	88	96	100	22	98	60	92	92	94	96	92	100

(The sample groups: G stands for the sound chestnut samples; S stands for the severely moldy chestnut samples. The preprocessing methods: NP = no preprocessing, VN = vector normalization, FD = first derivative, FD and VN = first derivative and vector normalization, SD = second derivative, and SD and VN = second derivative and vector normalization; the distance algorithm methods: SL = single linkage, CL = complete linkage, AL = average linkage, WAL = weighted average linkage, MA = median algorithm, CA = centroid algorithm) and WA = Ward's algorithm).

and 92%, respectively.

Selection of spectra bands

To determine the range of spectral bands that are reliable for accurate classification of moldy chestnuts, we compared results of different models based on the following band ranges: 833 - 2500, 1429 - 1648 and 1818 – 2085 nm.

The band of 1429 - 1648 and 1818 - 2085 nm are the sensitive bands of water, starch and proteins, the components vary conspicuously during the process of chestnut mildew. Therefore, the spectra in these bands would change accordingly. The models based on spectral data of sample group G and P, were calibrated by using the Ward's algorithm combined with four preprocessing methods (second derivative and vector normalization, second derivative, first derivative and vector normalization, and no preprocessing). Table 3 represented

Table 3. Calibration results based on different wavelength range.

Correct classification rate (%)		Wavelength range		
		833 - 2500 nm	1648 nm - 1429 m	1818 - 2085 nm
SD and W'A	G	93	62	74
	P	89	40	100
FD+VN and W'A	G	100	64	100
	P	92	20	98
SD+VN and W'A	G	90	100	45
	P	24	24	45
NP and W'A	G	69	24	41
	P	33	77	65

(The sample groups: G stands for the sound chestnut samples; P stands for the slightly moldy chestnut samples. The preprocessing methods: NP = no preprocessing, VN = vector normalization, FD = first derivative, FD and VN = first derivative and vector normalization, SD = second derivative, and SD and VN = second derivative and vector normalization; the distance algorithm methods: SL = single linkage, CL = complete linkage, AL = average linkage, WAL = weighted average linkage, MA = median algorithm, CA = centroid algorithm) and WA = Ward's algorithm).

Table 4. Validation results of the classification model of moldy chestnuts.

	Calibration set			Validation set		
	G	P	S	G	P	S
Correct classification rate (%)	100	98	100	100	92.8	100

(The sample groups: G stands for the sound chestnut samples; P stands for the slightly moldy chestnut samples; S stands for the severely moldy chestnut samples).

the results of the models based on different spectra bands.

It was found that the first derivative and vector normalization method was more effective and accurate, compared with the second derivative and vector normalization, second derivative and no preprocessing method. The poor performance of the model based on raw spectra was the result of noise in the spectra. The derivative was a powerful preprocessing method, in abstracting the useful information of spectra. However, this method involved the increase of noises when the order became higher. The problem of the second derivative was that, the noises were amplified too much to be corrected by the vector normalization. Furthermore, the spectral band of 1,818 - 2,085 nm was the most sensitive range for the classification model, as the average correct classification rates of the three models calibrated by the Ward's algorithm and the first derivative and vector normalization method were 96, 44 and 99%, respectively. This result indicated that the spectra in range of 1,429-1,648 nm had inadequate information for classi- fying the sample groups while the irrelative information in whole range spectra decreased the correct rate of the model.

Establishment and validation of the classification model

Based on the spectral data of the three chestnut groups in the band range of 1818 - 2085 nm, the classification model was established by using the first derivative and vector normalization for spectra preprocessing and the Ward's algorithm as the distance algorithm method. Table 4 showed the results from the classification model for identifying moldy chestnuts. It was found that the model was excellent at distinguishing the severe moldy chestnut from the sound chestnut.

The correct classification rate of slightly moldy chestnut was relatively lower. This could be attributed to the moldy part which was not shined by the spectra, so that the spectra only contained the information of the sound part of that sample. Another possible reason was that the algorithm method needed to be improved since the differences between sound chestnuts and slightly moldy ones were too tiny to be distinguished. However, although the correct classification rate of slightly moldy chestnut was not 100, 92.8% was still acceptable because of the absence of a method for detecting the slightly moldy chestnut in current produce.

Conclusion

A classification model based on the distance theory was established by using the NIR spectral data of intact chestnuts with different mildewing conditions. It was found that the characteristic information intermixed in the spectra of intact chestnut, can be purified by special preprocessing, such as first derivative and vector normalization, second derivative, etc. Furthermore, it was ensured that NIR spectroscopy is a reliable technology for classification of moldy chestnuts. By comparing the different effects, to distinguish distance algorithms, pre-processing methods and spectral band, it was found that the most accurate and effective model was established based on the spectral band of 1818-2085 nm by using the first derivative and vector normalization for spectra preprocessing and the Ward's algorithm, as distance algorithm method.

The classification model can be used to identify 100% sound chestnuts without mildew, 100% severe moldy chestnuts and 92.8% slightly moldy chestnuts. These results demonstrated that, the classification model based on near infrared spectral analyses, can be used to accurately detect moldy chestnuts, thereby providing a nondestructive means for identifying and removing moldy chestnuts with reduced labor compared with traditional methods. This is meaningful for the chestnut industry and also would be used as reference, for the measuring research of target with peel.

ACKNOWLEDGEMENTS

The authors really appreciate the financial support provided by Foundation of Ph.D Site in University (20090146110018) and China Scholarship Council (2008104710). Also, we gratefully acknowledge the help of Liu Pei during the spectra acquisition.

REFERENCES

Blakey RJ, Bower JP, Bertling I (2009). Influence of water and ABA supply on the ripening pattern of avocado (Persea americana Mill.) fruit and the prediction of water content using Near Infrared Spectroscopy. Postharvest Biol. Technol., 53 (1): 72-76.

Burks CS, Dowell FE, Xie F (2000). Measuring fig quality using near-infrared spectroscopy. J. Stored Prod. Res., 36: 289-296.

Chen QS, Zhao JW, Lin H (2009). Study on discrimination of Roast green tea (Camellia sinensis L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition. Spectrochim Acta A Mol. Biomol .Spectrosc., 72: 845-850.

Cocchi M, Corbellini M, Foca G, Lucisano M, Pagani MA, Tassi L, Ulrici A (2005). Classification of bread wheat flours in different quality categories by a wavelet-based feature selection/classification algorithm on NIR spectra. Anal. Chim. Acta., 544: 100-107.

Di Egidio V, Sinelli N, Limbo S, Torri L, Franzetti L, Casiraghi E (2009). Evaluation of shelf-life of fresh-cut pineapple using FT-NIR and FT-IR spectroscopy. Postharvest Biol. Technol., 54: 87-92.

Feng Y, Qin L, Li F (2007). Technique of chestnut cultivation, China agricultural university press, BeiJing.

Flores-Rojas K, Sanchez MT, Perez-Marin D, Guerrero JE, Garrido-Varo A (2009). Quantitative assessment of intact green asparagus quality by near infrared spectroscopy. Postharvest Biol. Technol., 52(3): 300-306.

Freinkel S (2009). American Chestnut: The Life, Death, and Rebirth of a Perfect Tree. University of California Press, Ewing.

Liu F, He Y (2008). Classification of brands of instant noodles using Vis/NIR spectroscopy and chemometrics. Food Res. Int., 41: 562-567.

Lu WZ, Yuan HF, Xu GT (2007). Morden near infrared spectroscopy analytical technology, China Petrochemical Press, Beijing.

Neethirajan S, Karunakaran C, Jayas DS, White NDG (2007). Detection techniques for stored-product insects in grain. Food Control., 18: 157-162.

Nicolai BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI, Lammertyn J (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biol . Technol., 46: 99-118.

Roggo Y, Duponchel L, Huvenne JP (2003). Comparison of supervised pattern recognition methods with McNemar's statistical test - Application to qualitative analysis of sugar beet by near-infrared spectroscopy. Anal. Chim. Acta., 477: 187-200.

Sundaram J, Kandala CVK, Butts CL, Windham WR (2009). Application of NIR Reflectance Spectroscopy on Determination of Moisture Content of Peanuts: A Non Destructive Analysis Method. Annual meeting of ASABE. pp. 095864.

Tewari JC, Dixit V, Cho BK, Malik KA (2008). Determination of origin and sugars of citrus fruits using genetic algorithm, correspondence analysis and partial least square combined with fiber optic NIR spectroscopy. Spectrochim. Acta, Part A 71 (3): 1119-1127.

Tripathi S, Mishra HN (2009). A rapid FT-NIR method for estimation of aflatoxin B-1 in red chili powder. Food Control., 20: 840-846.

Wang L, Lee FSC, Wang XR (2007). Near-infrared spectroscopy for classification of licorice (Glycyrrhizia uralensis Fisch) and prediction of the glycyrrhizic acid (GA) content. Lwt-Food Sci. Technol., 40: 83-88.

Yan TL, Zhao L, Han D (2007). Basement and application of near infrared spectroscopy analysis, China Light Industry Press, Beijing.